

Hi

In order to test hypothesis of the churn by the customer price sensitivity. We need following data to build the model effectively

Data needed:

1. Customer data – includes industry, historical electricity consumption, date joined as customer etc.
2. Churn data - if customer has churned or not
3. 3. Historical price data – customer price for both electricity and gas at granular time intervals

After getting the data we would do feature engineering to eliminate the unused features in the model building process. And build the model such as Logistic Regression, Random forest, Gradient Boosted, Naive Bayes classifier these are the model used classification purpose. Which model perform best on the 3 types of datasets. We can test about magnitude of churn price and other important relative information and later we can decide about the discounting strategy allowed by the client side.

Regards,